

On Stability of Multistage Stochastic Programs

CHRISTIAN KÜCHLER*

Humboldt–Universität zu Berlin, Unter den Linden 6,
D–10099 Berlin, Germany, ckuechler@math.hu-berlin.de

Abstract

We study quantitative stability of linear multistage stochastic programs under perturbations of the underlying stochastic processes. It is shown that the optimal values behave Lipschitz continuous with respect to an L_p -distance. Therefore, we have to make a crucial regularity assumption on the conditional distributions, that allows to establish continuity of the recourse function with respect to the current state of the stochastic process. The main stability result holds for nonanticipative discretizations of the underlying process and thus represents a rigorous justification of established discretization techniques.

AMS 2000 subject classification: 90C15, 90C31

Key words: stochastic programming, multistage, stability, probability metrics

Introduction

Many stochastic optimization problems of practical interest do not allow for an analytic solution. Numerical approaches require the underlying probability measure to have finite support, which should be at the most of moderate size. Whenever the initial probability measure does not meet these demands, it has to be approximated by an auxiliary measure. It is obvious that the optimal value and the set of optimal decisions of the auxiliary problem should be close to the initial ones. Consequently, perturbation and stability analysis of stochastic programs are necessary for the development of reliable techniques for discretization and scenario reduction. While stability properties are well understood for non-dynamic chance constrained and two-stage problems, cf. the recent survey of Römis (2003), it turned out that the multistage case is more intricate. Recently, the latter situation has been studied by a variety of authors and the following

*Support by the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) and by the Bundesministerium für Bildung und Forschung (BMBF) under the grant 03SF0312E is gratefully acknowledged.

references are not exhausting at all. Statistical bounds have been provided by Shapiro (2003). Pennanen (2005) established asymptotic stability of specific approximations for a general class of convex multistage problems in terms of epi-convergence and he noticed, that quantitative results would require stronger assumptions. Indeed, the restriction on models with continuous decisions allowed Mirkov and Pflug (2006) to establish such a quantitative stability result for their tree approximations. Heitsch, Römisch, and Strugarek (2006) abstained from regularity conditions on decisions and underlying processes and, consequently, their quantitative stability result, considering arbitrary perturbations of the underlying process, incorporates a term measuring the distance of the filtrations induced by the initial and the auxiliary process, respectively. Vanishing in the two-stage case, this term reflects the relevance of the information structure and of the nonanticipativity constraints for multistage decision problems. We refer also to Barty (2004) who studied the role of information in stochastic optimization problems and introduced and reviewed several concepts of distances between filtrations.

The recent approach of Heitsch and Römisch (2005) aims to incorporate filtration distances into the construction of scenario trees. However, this requires some extra effort and, to the best of our understanding, these distances are not taken into account by a variety of established techniques. Thus, the main purpose of this paper is to provide general conditions under which these somewhat delicate terms may be omitted.

A main difficulty seems to be that without additional assumptions neither the recourse function nor an optimal decision will depend continuously on the current state of the underlying process, in general. Rockafellar and Wets (1974) showed that under weak conditions the optimal value can be achieved by *continuous decisions*, asymptotically. However, while this allows to deduce convergence results as those of Pennanen (2005), it does not lead to quantitative estimates. For deriving *continuity of the recourse function* and bounds based on a barycentric approximation scheme, Kuhn (2005) required the underlying processes to be autoregressive. He also indicated, that *the key element in any scenario tree construction is the discretization of the conditional probabilities*. We agree and underline that, in particular, continuous dependency of these probabilities on the current state of the underlying process is necessary for potential continuity of the recourse function and can be seen as *continuity of the available information w.r.t. the current state*. It is illustrated by Example 2.6 of Heitsch, Römisch, and Strugarek (2006) that the latter property is indispensable in order to omit any filtration distances and to obtain a good approximation of the initial process by usual techniques, that are based on stagewise clustering. Thus, we ensure by Assumption 2.6 the Lipschitz continuity of the conditional distributions, which allows to verify the same property for the recourse function in Theorem 1. With this at hand, we estimate in Theorem 2 the gap between the optimal value and the costs of a decision that is locally *calm*. This leads to our main result, Theorem 3, that provides an upper bound for the perturbation of the optimal

value.

Acknowledgement. We are grateful to Prof. Werner Römisch for his help and encouragement and to Stefan Vigerske for being perpetually willing for helpful discussions.

Notation and Conventions. Random variables are denoted by bold letters, e.g. $\boldsymbol{\xi}$ or \mathbf{x} , in contrast to their realizations, i.e., elements of their support, which are denoted by ξ or x , respectively. The notation $\boldsymbol{\xi}^t$ is used for the vector (ξ_1, \dots, ξ_t) and $\|\cdot\|$ denotes the maximum norm on \mathbb{R}^n for the respective value $n \in \mathbb{N}$.

1 Problem Formulation

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ we consider an \mathbb{R}^s -valued stochastic process $\boldsymbol{\xi} = (\xi_t)_{t=1}^T$ with time horizon $T \in \mathbb{N}$ and the associated filtration $(\mathcal{F}_t)_{t=1}^T$ defined through $\mathcal{F}_t \triangleq \sigma(\boldsymbol{\xi}^t)$ for $t = 1, \dots, T$. We assume that $\mathcal{F}_1 = \{\Omega, \emptyset\}$, $\boldsymbol{\xi}^T \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for every $p \in [1, +\infty)$, and set

$$\mathbb{P}^t \triangleq \mathbb{P}[\boldsymbol{\xi}^t \in \cdot] \quad \text{and} \quad \Xi^t \triangleq \text{supp } \mathbb{P}^t \subset \mathbb{R}^{s \cdot t} \quad \text{for } t = 1, \dots, T.$$

Furthermore, we consider the *costs* $b_t(\cdot)$, the *technology matrices* $A_{t,1}(\cdot)$, and the *right-hand sides* $h_t(\cdot)$, which all are assumed to depend affinely linear on $\xi_t \in \Xi_t$, $t = 1, \dots, T$. Altogether, they define the set-valued mappings (or, multifunctions)

$$\begin{aligned} M_t : X_{t-1} \times \Xi_t &\rightrightarrows X_t, \\ M_t(x_{t-1}, \xi_t) &\triangleq \{x_t \in X_t : A_{t,0}x_t + A_{t,1}(\xi_t)x_{t-1} = h_t(\xi_t)\} \end{aligned}$$

for certain nonempty, closed, and polyhedral sets $X_t \subset \mathbb{R}^m$ and $t = 1, \dots, T$. The objective function is given by

$$\begin{aligned} \varphi : \mathbb{R}^{m \cdot T} \times \Xi^T &\rightarrow \mathbb{R} \cup \{\pm\infty\}, \\ \varphi(x_1, \dots, x_T, \xi^T) &\triangleq \begin{cases} \sum_{t=1}^T \langle b_t(\xi_t), x_t \rangle & \text{if } x_1 \in X_1, x_t \in M_t(x_{t-1}, \xi_t), \ t = 2, \dots, T \\ +\infty & \text{else.} \end{cases} \end{aligned}$$

A tuple $\mathbf{x} = (x_1, \dots, x_T)$ of Borel-measurable mappings $x_t : \Xi^t \rightarrow X_t$, $t = 1, \dots, T$, is called a *feasible decision* w.r.t. $\boldsymbol{\xi}$, if the recourse equation

$$(1) \quad \mathbf{x}_t(\boldsymbol{\xi}^t) \in M_t(\mathbf{x}_{t-1}(\boldsymbol{\xi}^{t-1}), \boldsymbol{\xi}_t)$$

is fulfilled \mathbb{P} -a.s. for $t = 2, \dots, T$. The class of feasible decisions \mathbf{x} will be denoted by $\mathcal{S}(\boldsymbol{\xi})$ and we set $x_0 = 1$ for the sake of notational convenience.

We want to study the following linear multistage optimization problem:

$$(2) \quad v(\boldsymbol{\xi}) \triangleq \min_{\mathbf{x} \in \mathcal{S}(\boldsymbol{\xi})} \mathbb{E}[\varphi(\mathbf{x}, \boldsymbol{\xi}^T)],$$

and to establish a bound for the perturbation of $v(\boldsymbol{\xi})$ when $\boldsymbol{\xi}$ is replaced by another process $\tilde{\boldsymbol{\xi}}$.

The polyhedral form of M_t allows to conclude from Rockafellar and Wets (1998)'s Example 9.35 that M_t is Lipschitz continuous in x_{t-1} and ξ_t with respect to the Pompeiu-Hausdorff distance \mathfrak{d} in the following sense. There exists a constant $M \geq 0$ with

$$\begin{aligned} \mathfrak{d}(M_t(x_{t-1}, \xi_t), M_t(\hat{x}_{t-1}, \xi_t)) &\leq M \cdot \max\{1, \|\xi_t\|\} \cdot \|\hat{x}_{t-1} - x_{t-1}\| \quad \text{and} \\ \mathfrak{d}(M_t(x_{t-1}, \xi_t), M_t(x_{t-1}, \hat{\xi}_t)) &\leq M \cdot \max\{1, \|x_{t-1}\|\} \cdot \|\hat{\xi}_t - \xi_t\|, \end{aligned}$$

for every $\xi_t, \hat{\xi}_t \in \Xi_t$ and $x_{t-1}, \hat{x}_{t-1} \in X_{t-1}$. We recall that the Pompeiu-Hausdorff distance between two sets $A, B \subset \mathbb{R}^m$ is defined by

$$\mathfrak{d}(A, B) \triangleq \max \left\{ \sup_{a \in A} \text{dist}(a, B), \sup_{b \in B} \text{dist}(b, A) \right\}.$$

Remark 1.1. *The Lipschitz continuity of M_t was our unique motivation to presume linear recourse. Analogously, this is true for the linear costs $\langle b_t(\xi_t), x_t \rangle$, where we use only the existence of a constant B with*

$$\begin{aligned} \|\langle b_t(\xi_t), x_t \rangle - \langle b_t(\hat{\xi}_t), x_t \rangle\| &\leq B \|\xi_t - \hat{\xi}_t\| \|x_t\| \quad \text{and} \\ \|\langle b_t(\xi_t), x_t \rangle - \langle b_t(\xi_t), \hat{x}_t \rangle\| &\leq B \max\{1, \|\xi_t\|\} \|x_t - \hat{x}_t\|. \end{aligned}$$

Furthermore, all results remain valid if M_t , h_t , and b_t depend on $\boldsymbol{\xi}^t$ instead of $\boldsymbol{\xi}_t$.

The integrability condition on $\boldsymbol{\xi}^T$ is due to notational simplicity. Actually, it suffices to have $\boldsymbol{\xi}^T \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for a sufficiently large $p \in \mathbb{R}_+$.

2 Continuity of the Recourse Function

Let $V_t : \Xi^t \times X_{t-1} \rightarrow \mathbb{R}$ be the recourse function at time t , i.e., $V_t(\xi^T, x_{t-1})$ represents the minimal achievable expected future costs after having chosen $\mathbf{x}_{t-1} = x_{t-1}$, having observed $\{\boldsymbol{\xi}^t = \xi^t\}$, and before deciding on x_t . It is defined recursively by $V_{T+1} \triangleq 0$ and the Dynamic Programming Equation

$$V_t(\xi^t, x_{t-1}) \triangleq \inf_{x_t \in M_t(x_{t-1}, \xi_t)} \langle b_t(\xi_t), x_t \rangle + \mathbb{E} [V_{t+1}(\boldsymbol{\xi}^{t+1}, x_t) \mid \boldsymbol{\xi}^t = \xi^t] \quad \text{for } t = T, \dots, 1.$$

It was proved by Evstigneev (1976) that V_t is well defined and measurable under the following

Assumption 2.1.

- (i) *There exists an integrable random variable Q such that $\varphi(x, \boldsymbol{\xi}^T) \geq Q$ holds \mathbb{P} -a.s. for every $x \in \mathbb{R}^{m \cdot T}$.*

(ii) For each $c \in \mathbb{R}$ the random level set $\{x \in \mathbb{R}^{m \cdot T} : \varphi(x, \boldsymbol{\xi}^T) \leq c\}$ is compact \mathbb{P} -a.s.

A decision $\mathbf{x} \in \mathcal{S}(\boldsymbol{\xi})$ is optimal if and only if the equality

$$(3) \quad V_t(\xi^t, \mathbf{x}_{t-1}(\xi^{t-1})) = \langle b_t(\xi_t), \mathbf{x}_t(\xi^t) \rangle + \mathbb{E} [V_{t+1}(\boldsymbol{\xi}^{t+1}, \mathbf{x}_t(\xi^t)) | \boldsymbol{\xi}^t = \xi^t]$$

holds for \mathbb{P}^t -almost every $\xi^t \in \Xi^t$ and $t = 1, \dots, T$. Moreover, for every Borel measurable mapping $\mathbf{x}_{t-1} : \Xi^{t-1} \rightarrow X_{t-1}$ there exists a measurable $\mathbf{x}_t : \Xi^t \rightarrow X_t$ such that

$$(4) \quad V_t(\xi^t, \mathbf{x}_{t-1}(\xi^{t-1})) = \langle b_t(\xi_t), \mathbf{x}_t(\xi^t) \rangle + \mathbb{E} [V_{t+1}(\boldsymbol{\xi}^{t+1}, \mathbf{x}_t(\xi^t)) | \boldsymbol{\xi}^t = \xi^t]$$

holds true for \mathbb{P}^t -almost every $\xi^t \in \Xi^t$. Actually, Evstigneev (1976)'s results allow a further formulation of (4), that is more general with regard to the \mathbb{P}^t -null sets on which (4) does not hold. Indeed, the following corollary is an immediate consequence of applying Evstigneev (1976)'s Lemma 4 within the proof of his Theorem 2:

Corollary 2.2. *There exists a $\mathcal{B}(X_{t-1}) \otimes \mathcal{B}(\Xi^t)$ -measurable mapping $\eta_t : X_{t-1} \times \Xi^t \rightarrow X_t$, such that*

$$V_t(\xi^t, x_{t-1}) = \langle b_t(\xi_t), \eta_t(x_{t-1}, \xi^t) \rangle + \mathbb{E} [V_{t+1}(\boldsymbol{\xi}^{t+1}, \eta_t(x_{t-1}, \xi^t)) | \boldsymbol{\xi}^t = \xi^t]$$

holds $\mathbb{P}^t(d\xi^t)$ -a.s. for all $x_{t-1} \in X_{t-1}$.

The following assumption ensures complete recourse and the existence of optimal decisions that are bounded in a certain sense:

Assumption 2.3. *There is a constant $L \geq 1$ such that for $t = 1, \dots, T$ and certain Borel sets $A^{t'} \subset \Xi^t$ with $\mathbb{P}^t[A^{t'}] = 1$, $t = 1, \dots, T$, the following property holds: For every Borel measurable mapping $\mathbf{x}_{t-1} : \Xi^{t-1} \rightarrow X_{t-1}$ there exists a measurable $\mathbf{x}_t : \Xi^t \rightarrow X_t$ such that $\mathbf{x}_t(\xi^t) \in M_t(\mathbf{x}_{t-1}(\xi^{t-1}), \xi_t)$, identity (4), and*

$$(5) \quad \|\mathbf{x}_t(\xi^t)\| \leq L \cdot \max \{1, \|\mathbf{x}_{t-1}(\xi^{t-1})\|\} \cdot \max \{1, \|\xi^t\|\}.$$

hold true for every $\xi^t \in A^{t'}$.

Remark 2.4. *Unfortunately, the existence of decisions which are bounded in the above sense may be hard to verify, in general. However, (5) holds true for every $x_t \in M_t(x_{t-1}, \xi_t)$ if X_t is bounded, or, more general, whenever the projection of X_t onto the kernel of the recourse matrix $A_{t,0}$ is bounded.*

The linear growth condition (5) can be relaxed to polynomial growth, then the growth rate in ξ^t of the Lipschitz constant in Theorem 1 and the subsequent results will change accordingly.

Under the Assumptions 2.1 and 2.3, we can restrict ourselves on decisions \mathbf{x} satisfying

$$(6) \quad \|\mathbf{x}_t(\boldsymbol{\xi}^t)\| \leq L^t \cdot \max\{1, \|\boldsymbol{\xi}^t\|\}^{t-1} \quad \mathbb{P} - \text{a.s.}, t = 1, \dots, T,$$

which will be denoted as *bounded* in the following. Indeed, a tuple $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_T)$ of mappings with (1), (3), and (6) can be constructed by recursion, and from Theorem 14.37 of Rockafellar and Wets (1998) it follows that every \mathbf{x}_t can be chosen measurable. Consequently, \mathbf{x} is an optimal decision.

It is well-known that V_t tends to show some smoothness w.r.t. x_{t-1} . We refer to Birge and Louveaux (1997) and Ruszczyński and Shapiro (2003) who derive convexity as well as piecewise linearity for the case of finite Ξ^T and to Kuhn (2005) who proved continuity under compactness assumptions on Ξ^T and X_1, \dots, X_T . Thus, the following Proposition can be seen as an adaption of these results to our Lipschitz continuous framework.

Proposition 2.5. *The recourse function V_t is Lipschitz continuous w.r.t. the decision x_{t-1} in the following sense. There exists a constant $\bar{M} > 0$ such that for $t = 1, \dots, T$ and \mathbb{P}^t -almost every $\xi^t \in \Xi^t$ the relation*

$$(7) \quad |V_t(\xi^t, x_{t-1}) - V_t(\xi^t, \hat{x}_{t-1})| \leq [V_t]_{Lip}^x(\xi^t) \cdot \|x_{t-1} - \hat{x}_{t-1}\|$$

holds true for every $x_{t-1}, \hat{x}_{t-1} \in X_{t-1}$ with a (random) Lipschitz constant $[V_t]_{Lip}^x(\xi^t)$ satisfying

$$(8) \quad [V_t]_{Lip}^x(\xi^t) \leq \bar{M} \cdot \mathbb{E} [\max\{1, \|\boldsymbol{\xi}^T\|\}^{2+T-t} \mid \boldsymbol{\xi}^t = \xi^t].$$

Proof. The assertion is true for $V_{T+1} \equiv 0$. Assume it is true also for $s = t+1, \dots, T$ with Lipschitz constants $[V_s]_{Lip}^x$ and, for instance, assume that the difference on the left side of (7) is negative. Then, due to (4), there exists an $\mathbf{x}_t^*(\xi^t) \in M_t(x_{t-1}, \xi_t)$, such that the left side of (7) coincides for \mathbb{P}^t -a.e. ξ^t with

$$(9) \quad \begin{aligned} & -\langle b_t(\xi_t), \mathbf{x}_t^*(\xi^t) \rangle - \mathbb{E} [V_{t+1}(\boldsymbol{\xi}^{t+1}, \mathbf{x}_t^*(\xi^t)) \mid \boldsymbol{\xi}^t = \xi^t] \\ & + \inf_{\hat{x}_t \in M_t(\hat{x}_{t-1}, \xi_t)} \{ \langle b_t(\xi_t), \hat{x}_t \rangle + \mathbb{E} [V_{t+1}(\boldsymbol{\xi}^{t+1}, \hat{x}_t) \mid \boldsymbol{\xi}^t = \xi^t] \}. \end{aligned}$$

Moreover, it follows from Corollary 2.2 that we may assume that the $\mathbb{P}^t(d\xi^t)$ -null sets on which this identity does not hold coincide for all $x_{t-1} \in X_{t-1}$. Due to Theorem 14.37 of Rockafellar and Wets (1998) we can choose a measurable $\hat{\mathbf{x}}_t^*$ with

$$\hat{\mathbf{x}}_t^*(\xi^t) \in \arg \min_{z \in M_t(\hat{x}_{t-1}, \xi_t)} \|z - \mathbf{x}_t^*(\xi^t)\|$$

to bound (9) from above by

$$|\langle b_t(\xi_t), \mathbf{x}_t^*(\xi^t) - \hat{\mathbf{x}}_t^*(\xi^t) \rangle| + \mathbb{E} [V_{t+1}(\boldsymbol{\xi}^{t+1}, x_t^*) - V_{t+1}(\boldsymbol{\xi}^{t+1}, \hat{x}_t^*) \mid \boldsymbol{\xi}^t = \xi^t].$$

Linear growth of b_t and Lipschitz continuity of V_{t+1} w.r.t. x_t entail that this term is not greater than

$$\left(B \max\{1, \|\xi_t\|\} + \mathbb{E} \left[[V_{t+1}]_{Lip}^x(\xi^{t+1}) \mid \xi^t = \xi^t \right] \right) \cdot \|\mathbf{x}_t^*(\xi^t) - \hat{\mathbf{x}}_t^*(\xi^t)\|,$$

again $\mathbb{P}^t(d\xi^t)$ -a.s. for every $x_{t-1}, \hat{x}_{t-1} \in X_{t-1}$. By definition of $\hat{\mathbf{x}}_t^*$ and Lipschitz continuity of M_t , the latter term is bounded from above by

$$\left(MB \max\{1, \|\xi_t\|^2\} + M \max\{1, \|\xi_t\|\} \cdot \mathbb{E} \left[[V_{t+1}]_{Lip}^x(\xi^{t+1}) \mid \xi^t = \xi^t \right] \right) \cdot \|x_{t-1} - \hat{x}_{t-1}\|.$$

An analogous estimate holds whenever the difference on the left side of (7) is positive. Hence, $[V_t]_{Lip}^x(\xi^t)$ is given by the term in parentheses, from where we conclude by recursion that we can put

$$[V_t]_{Lip}^x(\xi^t) \triangleq B \sum_{i=t}^T M^{i-t+1} \mathbb{E} \left[\max\{1, \|\xi_i\|^2\} \cdot \prod_{k=t}^{i-1} \max\{1, \|\xi_k\|\} \mid \xi^t = \xi^t \right].$$

Finally, the asserted bound for $[V_t]_{Lip}^x$ results from a straightforward estimate. \square

Establishing continuity of $\xi^t \mapsto V_t(\xi^t, x_{t-1})$ is more subtle, since, unlike the decision variable x_{t-1} , the state ξ^t impacts not only the Lipschitz continuous time coupling constraints at time t , but also the expectations about the uncertainty after time t . Therefore, one can hardly expect V_t to be Lipschitz continuous w.r.t. ξ^t without having that *the conditional distribution of $(\xi_s)_{s=t+1}^T$ under $\{\xi^t = \xi^t\}$ depends continuously on ξ^t* , in some sense. This is illustrated by Example 2.6. of Heitsch, Römisch, and Strugarek (2006). Thus, for establishing recursively the continuity of V_t , we need that continuity of V_{t+1} w.r.t. ξ^{t+1} is passed down to the mapping $\xi^t \mapsto \mathbb{E} [V_{t+1}(\xi^{t+1}, x_t) \mid \xi^t = \xi^t]$. To this end, we introduce for $p \geq 1$ and a given Borel set $A^{t+1} \subset \Xi^{t+1}$ with $\mathbb{P}^{t+1}[A^{t+1}] = 1$ the class

$$\mathbb{F}_p^{A^{t+1}}(\Xi^{t+1}) \triangleq \left\{ f : \Xi^{t+1} \rightarrow \mathbb{R} : (10) \text{ holds for } \xi^{t+1}, \hat{\xi}^{t+1} \in A^{t+1} \right\}$$

and the Lipschitz condition

$$(10) \quad |f(\xi^{t+1}) - f(\hat{\xi}^{t+1})| \leq \max\{1, \|\xi^{t+1}\|, \|\hat{\xi}^{t+1}\|\}^{p-1} \|\xi^{t+1} - \hat{\xi}^{t+1}\|.$$

We recall that - except for our disregarding of the \mathbb{P}^{t+1} -null set $\Xi^{t+1} \setminus A^{t+1}$ within the definition of $\mathbb{F}_p^{A^{t+1}}$ - the p -th order Fortet-Mourier distance between probability measures P, Q on Ξ^{t+1} is defined by

$$\zeta_p^{A^{t+1}}(P, Q) \triangleq \sup_{f \in \mathbb{F}_p^{A^{t+1}}(\Xi^{t+1})} \left| \int_{\Xi^{t+1}} f(\xi^{t+1}) P(d\xi^{t+1}) - \int_{\Xi^{t+1}} f(\xi^{t+1}) Q(d\xi^{t+1}) \right|,$$

see, e.g., Rachev (1991) and Römisch (2003). Using this notation, the claimed continuity of the conditional distributions is specified by the following

Assumption 2.6. *There exist constants $W, K > 0$ and $r \geq 1$, such that with*

$$(11) \quad m_t \triangleq 2 + (T - t)(r + 1) \quad \text{for } t = 1, \dots, T,$$

the following conditions are fulfilled.

- (i) *For every $t = 1, \dots, T - 1$, every Borel set $A^{t+1} \subset \Xi^{t+1}$ with $\mathbb{P}^{t+1}[A^{t+1}] = 1$, and \mathbb{P}^t -a.e. $\xi^t, \hat{\xi}^t \in \Xi^t$*

$$\begin{aligned} & \zeta_{m_{t+1}}^{A^{t+1}} \left(\mathbb{P} [\xi^{t+1} \in \cdot \mid \xi^t = \xi^t], \mathbb{P} [\xi^{t+1} \in \cdot \mid \xi^t = \hat{\xi}^t] \right) \\ & \leq K \max \left\{ 1, \|\xi^t\|, \|\hat{\xi}^t\| \right\}^{m_t-1} \|\xi^t - \hat{\xi}^t\|. \end{aligned}$$

- (ii) *For every $t = 1, \dots, T - 1$ and \mathbb{P}^t -a.e. $\xi^t \in \Xi^t$*

$$\mathbb{E} [\max\{1, \|\xi^T\|\}^{2+T-t} \mid \xi^t = \xi^t] \leq W \cdot \max\{1, \|\xi^t\|\}^{m_t}.$$

Since the above assumption is crucial for the following continuity and stability results, it is discussed by the following

Remark 2.7. *Condition (i) is related to terms usually related to Markov processes, namely the coefficient of ergodicity and the Feller property, see, e.g. Dobrushin (1956) and Dynkin (1965), respectively. A similar assumption has been made by Bally, Pagès, and Printems (2005) to ensure stability of an optimal-stopping problem in a Markovian framework and by Mirkov and Pflug (2006) for their study of consistency of tree approximations. It is also made implicitly by Kuhn (2005) by focusing on autoregressive processes. The more involved formulation of Assumption 2.6, allowing for polynomially growing Lipschitz constants, is due to the fact that neither $\langle b_t(\xi_t), x_t \rangle$ nor M_{t+1} are uniformly Lipschitz continuous in ξ_t and x_t , unless both the support Ξ^T and the sets $X_t, t = 1, \dots, T$, are bounded. Indeed, under such a boundedness condition (i) may be significantly simplified.*

Lemma A.1 in the Appendix provides conditions on ξ , under which both (i) and (ii) hold true. In particular, this is the case if Ξ^T is finite. Then $\zeta_{m_{t+1}}$ is the optimal value of a linear optimization problem that can be solved numerically to determine the constants K and r .

We indicate that the definition of m_t allows the growth rate of the Lipschitz constant of $\xi^t \mapsto \mathbb{E}[f(\xi^{t+1}) \mid \xi^t = \xi^t]$ to exceed those of f 's Lipschitz constant by the value $r \geq 1$.

The following Theorem shows that Assumption 2.6 provides indeed Lipschitz continuity of V_t w.r.t. ξ^t . We refer to Proposition 2.7 of Kuhn (2005), which represents a corresponding continuity result.

Theorem 1. Suppose the Assumptions 2.1, 2.3, and 2.6 are fulfilled. For every $t = 1, \dots, T$ there is a constant $C_t > 0$ and a Borel set $A^t \subset \Xi^t$ with $\mathbb{P}^t[A^t] = 1$ such that

$$\frac{1}{C_t \max\{1, \|x_{t-1}\|\}} V_t(\cdot, x_{t-1}) \in \mathbb{F}_{m_t+1}^{A^t}(\Xi^t),$$

holds true for every $x_{t-1} \in X_{t-1}$.

Proof. The assertion holds true for $V_{T+1} \equiv 0$, we show that it follows recursively for $t \leq T$. To this end, we proceed as in the proof of Proposition 2.5 and choose a measurable \mathbf{x}_t^* with $\mathbf{x}_t^*(\xi^t) \in M_t(x_{t-1}, \xi_t)$, that fulfills (4) and $\|\mathbf{x}_t^*(\xi^t)\| \leq L \cdot \max\{1, \|x_{t-1}\|\} \cdot \max\{1, \|\xi^t\|\}$. Thus, we obtain

$$(12) \quad \begin{aligned} & |V_t(\xi^t, x_{t-1}) - V_t(\hat{\xi}^t, x_{t-1})| \\ &= |\langle b_t(\xi_t), \mathbf{x}_t^*(\xi^t) \rangle + \mathbb{E}[V_{t+1}(\boldsymbol{\xi}^{t+1}, \mathbf{x}_t^*(\xi^t)) | \boldsymbol{\xi}^t = \xi^t] \\ &\quad - \inf_{\hat{x}_t \in M_t(x_{t-1}, \hat{\xi}_t)} \left\{ \langle b_t(\hat{\xi}_t), \hat{x}_t \rangle + \mathbb{E}[V_{t+1}(\boldsymbol{\xi}^{t+1}, \hat{x}_t) | \boldsymbol{\xi}^t = \hat{\xi}^t] \right\}|, \end{aligned}$$

which holds, due to Assumption 2.3, for every $\xi^t, \hat{\xi}^t$ in a $\mathbb{P}^t - 1$ set $A^{t'}$ for all $x_{t-1} \in X_{t-1}$. We consider the case when the term under the norm is negative and choose a measurable $\hat{\mathbf{x}}_t^*$ with

$$\hat{\mathbf{x}}_t^*(\hat{\xi}^t) \in \operatorname{argmin}_{z \in M_t(x_{t-1}, \hat{\xi}_t)} \|z - \mathbf{x}_t^*(\xi^t)\|,$$

to obtain the following upper bound for (12):

$$(13) \quad \begin{aligned} & -\langle b_t(\xi_t), \mathbf{x}_t^*(\xi^t) \rangle - \mathbb{E}[V_{t+1}(\boldsymbol{\xi}^{t+1}, \mathbf{x}_t^*(\xi^t)) | \boldsymbol{\xi}^t = \xi^t] \\ & + \langle b_t(\hat{\xi}_t), \hat{\mathbf{x}}_t^*(\hat{\xi}^t) \rangle + \mathbb{E}[V_{t+1}(\boldsymbol{\xi}^{t+1}, \hat{\mathbf{x}}_t^*(\hat{\xi}^t)) | \boldsymbol{\xi}^t = \hat{\xi}^t]. \end{aligned}$$

Using linearity of b_t and Lipschitz continuity of M_t , the difference of the $\langle \cdot, \cdot \rangle$ -terms can be estimated by

$$(14) \quad \begin{aligned} & |\langle b_t(\xi_t), \mathbf{x}_t^*(\xi^t) \rangle - \langle b_t(\hat{\xi}_t), \mathbf{x}_t^*(\xi^t) \rangle| + |\langle b_t(\hat{\xi}_t), \mathbf{x}_t^*(\xi^t) \rangle - \langle b_t(\hat{\xi}_t), \hat{\mathbf{x}}_t^*(\hat{\xi}^t) \rangle| \\ & \leq B \|\xi_t - \hat{\xi}_t\| \cdot L \cdot \max\{1, \|x_{t-1}\|\} \max\{1, \|\xi^t\|\} \\ & \quad + B \max\{1, \|\hat{\xi}^t\|\} \cdot M \max\{1, \|x_{t-1}\|\} \|\xi_t - \hat{\xi}_t\| \\ & \leq B(L + M) \max\{1, \|x_{t-1}\|\} \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\} \|\xi_t - \hat{\xi}_t\|, \end{aligned}$$

The difference of the conditional expectations in (13) is bounded by

$$\begin{aligned} & |\mathbb{E}[V_{t+1}(\boldsymbol{\xi}^{t+1}, \mathbf{x}_t^*(\xi^t)) | \boldsymbol{\xi}^t = \xi^t] - \mathbb{E}[V_{t+1}(\boldsymbol{\xi}^{t+1}, \mathbf{x}_t^*(\xi^t)) | \boldsymbol{\xi}^t = \hat{\xi}^t]| \\ & + |\mathbb{E}[V_{t+1}(\boldsymbol{\xi}^{t+1}, \mathbf{x}_t^*(\xi^t)) | \boldsymbol{\xi}^t = \hat{\xi}^t] - \mathbb{E}[V_{t+1}(\boldsymbol{\xi}^{t+1}, \hat{\mathbf{x}}_t^*(\hat{\xi}^t)) | \boldsymbol{\xi}^t = \hat{\xi}^t]| \\ & \leq KC_{t+1} \max\{1, \|\mathbf{x}_t^*(\xi^t)\|\} \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^{m_t-1} \|\xi^t - \hat{\xi}^t\| \\ & \quad + \mathbb{E}[V_{t+1}^x(\boldsymbol{\xi}^{t+1}) | \boldsymbol{\xi}^t = \hat{\xi}^t] \cdot M \max\{1, \|x_{t-1}\|\} \cdot \|\xi^t - \hat{\xi}^t\|, \end{aligned}$$

whereby the last inequality follows on the one hand from the assertion for V_{t+1} and Assumption 2.6, on the other hand from Proposition 2.5, and the Lipschitz continuity of M_t . This estimate holds true for every $\xi^t, \hat{\xi}^t \in A^{t''} \cap A^{t'''}$ for all $x_{t-1} \in X_{t-1}$, where $A^{t''}$ and $A^{t'''}$ denote the $\mathbb{P}^t - 1$ -sets, on which the assertions of Proposition 2.5 and Assumption 2.6 hold, respectively.

Applying the estimate (8) as well as condition (ii) of Assumption 2.6 and using the boundedness of $\|\mathbf{x}_t^*\|$, we get that the latter sum is again bounded from above by

$$(15) \quad (KC_{t+1}L + \bar{M}WM) \max\{1, \|x_{t-1}\|\} \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^{m_t} \cdot \|\xi^t - \hat{\xi}^t\|.$$

The upper bounds (14) and (15) remain valid if the term under the norm in (12) is positive. Piecing all together, the assertion for V_t follows with $A^t \triangleq A^{t'} \cap A^{t''} \cap A^{t'''}$ and the Lipschitz constant C_t can be chosen by collecting the constants from (14) and (15), i.e.,

$$C_t \triangleq B(M + L) + KC_{t+1}L + \bar{M}WM.$$

□

Whenever an auxiliary process $\tilde{\xi}$ is expected to approximate ξ with regard to the optimization problem (2), it is indispensable that $\tilde{\xi}$ is nonanticipative w.r.t. ξ . This is illustrated, for the sake of completeness, by Example A.3 in the Appendix. Nonanticipativity is ensured in the following by

Definition 2.8. *A stochastic process $\tilde{\xi}$ on $(\Omega, \mathcal{F}, \mathbb{P})$ is called a discretization of ξ , if there exist Borel-measurable mappings*

$$f_t : \Xi^t \rightarrow \Xi_t \quad t = 1, \dots, T,$$

fulfilling the following conditions:

- (i) $\tilde{\xi}_t = f_t(\xi^t)$ for $t = 1, \dots, T$,
- (ii) for every $\xi^T \in f^T(\Xi^T)$ we have $f^T(\xi^T) = \xi^T$,
- (iii) $f_1 = \text{Id}$, and
- (iv) $f^T(\xi^T) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for every $p \in [1, \infty)$.

Thereby, $f^t(\xi^t)$ denotes the vector $(f_i(\xi^i))_{i=1}^t \in \mathbb{R}^{s^t}$, for $t = 2, \dots, T$.

Remark 2.9. *The nonanticipativity condition (i) is equivalent to $\sigma(\xi^t)$ -measurability of the random variable $\tilde{\xi}_t$. Condition (ii) is fulfilled, e.g., if f^T is the projection onto the set $f^T(\Xi^T)$. It is needed in the following sections for the identity (23) to hold. The integrability condition (iv) is assumed again for the sake of simplicity. For the following results, it suffices that $f^T(\xi^T) \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ for a constant $p \in \mathbb{R}_+$ that is sufficiently large.*

The following proposition relies heavily on the continuity of the recourse function stated in Theorem 1. It is shown that, although an optimal decision is not continuous, in general, its expected costs can be approximated by a piecewise constant decision. The latter may be infeasible, but it can be used to construct a feasible decision. This will be completed in the next section.

Proposition 2.10. *Consider an optimal decision \mathbf{x}^* that is bounded in the sense of (6) and a discretization mapping f^T according to Definition 2.8. There exists a constant $D > 0$ such that the following estimate holds:*

$$\left| v(\boldsymbol{\xi}) - \mathbb{E} \left[\sum_{t=1}^T \langle b_t(\boldsymbol{\xi}_t), \mathbf{x}_t^*(f^t(\boldsymbol{\xi}^t)) \rangle \right] \right| \leq D \mathbb{E} [\max\{1, \|\boldsymbol{\xi}^T\|, \|f^T(\boldsymbol{\xi}^T)\|\}^{m_1} \cdot \|\boldsymbol{\xi}^T - f^T(\boldsymbol{\xi}^T)\|],$$

where m_1 is defined by (11).

Proof. Due to $f_1 = Id$, we have to bound

$$\left| \mathbb{E} \left[\sum_{t=2}^T \langle b_t(\boldsymbol{\xi}_t), \mathbf{x}_t^*(\boldsymbol{\xi}^t) \rangle \right] - \mathbb{E} \left[\sum_{t=2}^T \langle b_t(\boldsymbol{\xi}_t), \mathbf{x}_t^*(f^t(\boldsymbol{\xi}^t)) \rangle \right] \right|.$$

By optimality of \mathbf{x}^* , the first sum is equal to $\mathbb{E} [V_2(\boldsymbol{\xi}^2, \mathbf{x}_1^*)]$ and it follows from Theorem 1 and boundedness of \mathbf{x}_1^* (and $\mathbf{x}_0^* \triangleq 1$) that

$$(16) \quad \mathbb{E} [|V_2(\boldsymbol{\xi}^2, \mathbf{x}_1^*) - V_2(f^2(\boldsymbol{\xi}^2), \mathbf{x}_1^*)|] \leq LC_2 \mathbb{E} [\max\{1, \|\boldsymbol{\xi}^2\|, \|f^2(\boldsymbol{\xi}^2)\|\}^{m_2} \cdot \|\boldsymbol{\xi}^2 - f^2(\boldsymbol{\xi}^2)\|]$$

Thus, it remains to bound

$$(17) \quad \left| \mathbb{E} \left[V_2(f^2(\boldsymbol{\xi}^2), \mathbf{x}_1^*) - \sum_{t=2}^T \langle b_t(\boldsymbol{\xi}_t), \mathbf{x}_t^*(f^t(\boldsymbol{\xi}^t)) \rangle \right] \right|.$$

To this end, we consider the following inequality

$$(18) \quad \left| \mathbb{E} \left[V_2(f^2(\boldsymbol{\xi}^2), \mathbf{x}_1^*) - \sum_{s=2}^{t-1} \langle b_s(\boldsymbol{\xi}_s), \mathbf{x}_s^*(f^s(\boldsymbol{\xi}^s)) \rangle - V_t(f^t(\boldsymbol{\xi}^t), \mathbf{x}_{t-1}^*(f^{t-1}(\boldsymbol{\xi}^{t-1}))) \right] \right| \leq D_t,$$

whose left side coincides with (17) for $t = T + 1$. It holds trivially for $t = 2$ with $D_2 = 0$ and we assume that it is also true for a certain $t \in \{2, \dots, T\}$ and some $D_t \geq 0$. To prove it recursively for $t + 1$, we aim to bound

$$(19) \quad \left| \mathbb{E} [V_t(f^t(\boldsymbol{\xi}^t), \mathbf{x}_{t-1}^*(f^{t-1}(\boldsymbol{\xi}^{t-1}))) - \langle b_t(\boldsymbol{\xi}_t), \mathbf{x}_t^*(f^t(\boldsymbol{\xi}^t)) \rangle - V_{t+1}(f^{t+1}(\boldsymbol{\xi}^{t+1}), \mathbf{x}_t^*(f^t(\boldsymbol{\xi}^t)))] \right|.$$

To this end, we use again \mathbf{x}^* 's optimality to expand the first summand:

$$(20) \quad \begin{aligned} & \mathbb{E}[V_t(f^t(\boldsymbol{\xi}^t), \mathbf{x}_{t-1}^*(f^{t-1}(\boldsymbol{\xi}^{t-1}))) \\ &= \int \langle b_t(f_t(\boldsymbol{\xi}^t)), \mathbf{x}_t^*(f^t(\boldsymbol{\xi}^t)) \rangle + \mathbb{E} [V_{t+1}(\boldsymbol{\xi}^{t+1}, \mathbf{x}_t^*(f^t(\boldsymbol{\xi}^t))) | \boldsymbol{\xi}^t = f^t(\boldsymbol{\xi}^t)] \mathbb{P}^t(d\boldsymbol{\xi}^t). \end{aligned}$$

Now, to estimate (19), we have to replace $b_t(f_t(\xi^t))$ by $b_t(\xi^t)$. Indeed, Lipschitz continuity of $b_t(\cdot)$ implies

$$|\langle b_t(f_t(\xi^t)), \mathbf{x}_t^*(f_t(\xi^t)) \rangle - \langle b_t(\xi_t), \mathbf{x}_t^*(f_t(\xi^t)) \rangle| \leq B \cdot \|\mathbf{x}_t^*(f_t(\xi^t))\| \cdot \|\xi^t - f_t(\xi^t)\|.$$

To estimate the difference of the V_{t+1} -terms in (19) and (20), we add and subtract the term $\mathbb{E} [V_{t+1}(f^{t+1}(\xi^{t+1}), \mathbf{x}_t^*(f_t(\xi^t))) | \xi^t = \xi^t]$ and use the triangle inequality to estimate

$$\begin{aligned} & |\mathbb{E} [V_{t+1}(\xi^{t+1}, \mathbf{x}_t^*(f_t(\xi^t))) | \xi^t = f^t(\xi^t)] - \mathbb{E} [V_{t+1}(f^{t+1}(\xi^{t+1}), \mathbf{x}_t^*(f_t(\xi^t))) | \xi^t = \xi^t]| \\ & \leq |\mathbb{E} [V_{t+1}(\xi^{t+1}, \mathbf{x}_t^*(f_t(\xi^t))) | \xi^t = f^t(\xi^t)] - \mathbb{E} [V_{t+1}(\xi^{t+1}, \mathbf{x}_t^*(f_t(\xi^t))) | \xi^t = \xi^t]| \\ & \quad + |\mathbb{E} [V_{t+1}(\xi^{t+1}, \mathbf{x}_t^*(f_t(\xi^t))) | \xi^t = \xi^t] - \mathbb{E} [V_{t+1}(f^{t+1}(\xi^{t+1}), \mathbf{x}_t^*(f_t(\xi^t))) | \xi^t = \xi^t]|. \end{aligned}$$

By applying Theorem 1 and Assumption 2.6 we conclude that this term is bounded for \mathbb{P}^t -almost every ξ^t by

$$\begin{aligned} & KC_{t+1} \max \{1, \|\mathbf{x}_t^*(f_t(\xi^t))\|\} \max \{1, \|\xi^t\|, \|f^t(\xi^t)\|\}^{m_t-1} \|\xi^t - f^t(\xi^t)\| \\ & + C_{t+1} \max \{1, \|\mathbf{x}_t^*(f_t(\xi^t))\|\} \\ & \cdot \mathbb{E} [\max \{1, \|\xi^{t+1}\|, \|f^{t+1}(\xi^{t+1})\|\}^{m_{t+1}} \|\xi^{t+1} - f^{t+1}(\xi^{t+1})\| | \xi^t = \xi^t] \\ & \leq (K+1)C_{t+1}L^t \mathbb{E} [\max \{1, \|\xi^{t+1}\|, \|f^{t+1}(\xi^{t+1})\|\}^{m_t+t-1} \|\xi^{t+1} - f^{t+1}(\xi^{t+1})\| | \xi^t = \xi^t], \end{aligned}$$

where the inequality follows from boundedness of \mathbf{x}_t^* and the relation $m_{t+1} \leq m_t - 1$. Integration w.r.t. $\mathbb{P}^t(d\xi^t)$ and combining these estimates with (20) entails that (19) does not exceed

$$(21) \quad (B + (K+1)C_{t+1})L^t \mathbb{E} [\max \{1, \|\xi^{t+1}\|, \|f^{t+1}(\xi^{t+1})\|\}^{m_t+t-1} \cdot \|\xi^{t+1} - f^{t+1}(\xi^{t+1})\|],$$

Hence, (18) holds for $t+1$ with D_{t+1} being equal to the sum of D_t and (21).

Due to the fact that both $m_t + t - 1$ and m_2 are smaller than m_1 , the sum of (16) and (17) does not exceed

$$D \mathbb{E} [\max \{1, \|\xi^T\|, \|f^T(\xi^T)\|\}^{m_1} \cdot \|\xi^T - f^T(\xi^T)\|]$$

with $D \triangleq LC_2 + D_{T+1}$. This completes the proof. \square

3 Calmness of Decisions

Whenever ξ is replaced by another process $\tilde{\xi}$, the perturbation of the optimal value in (2) can be estimated by considering the terms

$$\|\langle b_t(\xi_t), \mathbf{x}_t^*(\xi^t) \rangle - \langle b_t(\tilde{\xi}_t), \tilde{\mathbf{x}}_t^*(\tilde{\xi}^t) \rangle\| \quad \text{for } t = 1, \dots, T,$$

where \mathbf{x}^* and $\tilde{\mathbf{x}}^*$ are optimal decisions w.r.t. ξ and $\tilde{\xi}$, respectively. While the coefficients b_t are close whenever ξ is well approximated by $\tilde{\xi}$, this is not necessarily true for the

decisions. Indeed, $\tilde{\mathbf{x}}^*$ is constant on every set $\{\xi^t : f^t(\xi^t) = z\}, z \in f^t(\Xi^t)$, whereas \mathbf{x}^* does not depend continuously on ξ^t , in general. Hence, our next purpose is to point out for every bounded optimal decision the existence of a *calm* decision that generates similar expected costs. To this end, we consider an optimal decision \mathbf{x}^* which is bounded in the sense of (6) and set

$$\bar{\mathbf{x}}_1^* \triangleq \mathbf{x}_1^* \quad \text{and} \quad \bar{\mathbf{x}}_t^*(\xi^t) \in \operatorname{argmin}_{z \in M_t(\bar{\mathbf{x}}_{t-1}^*(\xi^{t-1}), \xi_t)} \|\mathbf{x}_t^*(f^t(\xi^t)) - z\| \quad \text{for } t = 2, \dots, T,$$

where, again due to Theorem 14.37 of Rockafellar and Wets (1998), the latter mappings can be chosen to be measurable. Due to Lipschitz continuity of M_t , the local variability of $\bar{\mathbf{x}}_t^*(\cdot)$ in ξ^t can be bounded recursively. More precisely, $\bar{\mathbf{x}}_t^*(\cdot)$ is *calm locally around* $f(\xi^t)$ for every $\xi^t \in \Xi^t$:

Proposition 3.1. *For every $t = 1, \dots, T$ and $\xi^t \in \Xi^t$ we have*

$$(22) \quad \|\bar{\mathbf{x}}_t^*(\xi^t) - \bar{\mathbf{x}}_t^*(f^t(\xi^t))\| \leq (T-1)M^{T-1} \max\{1, \|\xi^T\|, \|f^T(\xi^T)\|\}^{T-1} \|\xi^T - f^T(\xi^T)\|.$$

Proof. For $t = 1$, the difference on the left side of (22) vanishes. For $t > 1$ we use the identity

$$(23) \quad \bar{\mathbf{x}}_t^*(f^t(\xi^t)) = \mathbf{x}_t^*(f^t(\xi^t))$$

and the definition of $\bar{\mathbf{x}}_t^*(\xi^t)$ to write

$$\|\bar{\mathbf{x}}_t^*(\xi^t) - \bar{\mathbf{x}}_t^*(f^t(\xi^t))\| = \inf_{z \in M_t(\bar{\mathbf{x}}_{t-1}^*(\xi^{t-1}), \xi_t)} \|z - \bar{\mathbf{x}}_t^*(f^t(\xi^t))\|.$$

Using the triangle inequality as well as $\bar{\mathbf{x}}_t^*(f^t(\xi^t)) \in M_t(\bar{\mathbf{x}}_{t-1}^*(f^{t-1}(\xi^{t-1})), f_t(\xi^t))$ and Lipschitz continuity of M_t , this term can be estimated against

$$\begin{aligned} & M \max\{1, \|\xi^t\|\} \|\bar{\mathbf{x}}_{t-1}^*(\xi^{t-1}) - \bar{\mathbf{x}}_{t-1}^*(f^{t-1}(\xi^{t-1}))\| \\ & + M \max\{1, \|\bar{\mathbf{x}}_{t-1}^*(f^{t-1}(\xi^{t-1}))\|\} \|f^t(\xi^t) - \xi^t\|, \end{aligned}$$

and, by boundedness of \mathbf{x}_{t-1} , against

$$\begin{aligned} & M \max\{1, \|\xi^t\|\} \|\bar{\mathbf{x}}_{t-1}^*(\xi^{t-1}) - \bar{\mathbf{x}}_{t-1}^*(f^{t-1}(\xi^{t-1}))\| \\ & + ML \max\{1, \|f^{t-1}(\xi^{t-1})\|\}^{t-1} \|f^t(\xi^t) - \xi^t\|. \end{aligned}$$

Recursively, we obtain that the left side of (22) is bounded by

$$L \sum_{i=2}^t M^{t+1-i} \max\{1, \|f^{i-1}(\xi^{i-1})\|\}^{i-1} \max\{1, \|\xi^t\|\}^{t-i} \|\xi^i - f^i(\xi^i)\|.$$

The assertion follows by a straightforward estimate. \square

From now on, $\bar{\mathbf{x}}^*$ is referred to as \mathbf{x}^* 's *calm modification*.

The following theorem shows that the difference of the expected costs generated by \mathbf{x}^* and $\bar{\mathbf{x}}^*$ can be estimated in terms of the deviation between $\boldsymbol{\xi}^T$ and $f^T(\boldsymbol{\xi}^T)$.

Theorem 2. *Suppose the Assumptions 2.1, 2.3, and 2.6 are fulfilled. Consider an optimal decision \mathbf{x}^* which is bounded in the sense of (6) and its calm modification $\bar{\mathbf{x}}^*$. There exists a constant $C > 0$ such that the following estimate holds*

$$|\mathbb{E}\varphi(\mathbf{x}^*, \boldsymbol{\xi}^T) - \mathbb{E}\varphi(\bar{\mathbf{x}}^*, \boldsymbol{\xi}^T)| \leq C \mathbb{E} [\max\{1, \|\boldsymbol{\xi}^T\|, \|f^T(\boldsymbol{\xi}^T)\|\}^{m_1} \cdot \|\boldsymbol{\xi}^T - f^T(\boldsymbol{\xi}^T)\|].$$

Proof. As in the proof of Proposition 2.10, we have to estimate

$$\left| \mathbb{E} \left[\sum_{t=2}^T \langle b_t(\boldsymbol{\xi}_t), \mathbf{x}_t^*(\boldsymbol{\xi}^t) \rangle - \sum_{t=2}^T \langle b_t(\boldsymbol{\xi}_t), \bar{\mathbf{x}}_t^*(\boldsymbol{\xi}^t) \rangle \right] \right|$$

We apply Proposition 2.10 and the triangle inequality to replace in the first sum $\mathbf{x}_t^*(\boldsymbol{\xi}^t)$ by $\mathbf{x}_t^*(f^t(\boldsymbol{\xi}^t))$ and to estimate the resulting error. Then, it remains to bound

$$\left| \mathbb{E} \left[\sum_{t=2}^T \langle b_t(\boldsymbol{\xi}_t), \mathbf{x}_t^*(f^t(\boldsymbol{\xi}^t)) \rangle - \sum_{t=2}^T \langle b_t(\boldsymbol{\xi}_t), \bar{\mathbf{x}}_t^*(\boldsymbol{\xi}^t) \rangle \right] \right|.$$

The following upper bound is obtained by applying identity (23) as well as the calmness property of $\bar{\mathbf{x}}^*$ from Proposition 3.1:

$$\begin{aligned} & \mathbb{E} \left[\sum_{t=2}^T B \max\{1, \|\boldsymbol{\xi}^t\|\} (T-1) M^{T-1} \max\{1, \|\boldsymbol{\xi}^T\|, \|f^T(\boldsymbol{\xi}^T)\|\}^{T-1} \|\boldsymbol{\xi}^T - f^T(\boldsymbol{\xi}^T)\| \right] \\ & \leq (T-1)^2 B M^{T-1} \mathbb{E} [\max\{1, \|\boldsymbol{\xi}^T\|, \|f^T(\boldsymbol{\xi}^T)\|\}^T \|\boldsymbol{\xi}^T - f^T(\boldsymbol{\xi}^T)\|]. \end{aligned}$$

Finally, the sum of the latter term and the bound obtained from Proposition 2.10 is smaller than

$$C \mathbb{E} [\max\{1, \|\boldsymbol{\xi}^T\|, \|f^T(\boldsymbol{\xi}^T)\|\}^{m_1} \cdot \|\boldsymbol{\xi}^T - f^T(\boldsymbol{\xi}^T)\|],$$

with a constant $C \triangleq D + (T-1)^2 B M^{T-1}$. □

4 Stability

With the above results for calm decisions, we are ready to adress the question of stability.

Theorem 3. *Suppose the Assumptions 2.1, 2.3, and 2.6 are fulfilled and let $\tilde{\boldsymbol{\xi}}$ be a discretization of $\boldsymbol{\xi}$ according to Definition 2.8, that fulfills Assumption 2.3, too. There exists a constant $\gamma > 0$, such that*

$$|v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}})| \leq \gamma \mathbb{E} [\max\{1, \|\boldsymbol{\xi}^T\|, \|\tilde{\boldsymbol{\xi}}^T\|\}^{m_1} \cdot \|\boldsymbol{\xi}^T - \tilde{\boldsymbol{\xi}}^T\|]$$

holds.

Proof. To this end, we consider a bounded optimal decision $\mathbf{x}^* \in \mathcal{S}(\boldsymbol{\xi})$ and its calm modification $\bar{\mathbf{x}}^*$ from Section 3. Applying Theorem 2 yields the following inequality

$$(24) \quad \begin{aligned} v(\tilde{\boldsymbol{\xi}}) - v(\boldsymbol{\xi}) &= v(\tilde{\boldsymbol{\xi}}) - \mathbb{E}\varphi(\mathbf{x}^*, \boldsymbol{\xi}) \\ &\leq v(\tilde{\boldsymbol{\xi}}) - \mathbb{E}\varphi(\bar{\mathbf{x}}^*, \boldsymbol{\xi}) + C\mathbb{E} \left[\max\{1, \|\boldsymbol{\xi}^T\|, \|\tilde{\boldsymbol{\xi}}^T\|\}^{m_1} \cdot \|\boldsymbol{\xi}^T - \tilde{\boldsymbol{\xi}}^T\| \right]. \end{aligned}$$

Since the restriction of $\bar{\mathbf{x}}^*$ on $f^T(\Xi^T)$ is contained in $\mathcal{S}(\tilde{\boldsymbol{\xi}})$, we can write

$$(25) \quad \begin{aligned} v(\tilde{\boldsymbol{\xi}}) - \mathbb{E}\varphi(\bar{\mathbf{x}}^*, \boldsymbol{\xi}) &\leq \mathbb{E}\varphi(\bar{\mathbf{x}}^*, \tilde{\boldsymbol{\xi}}) - \mathbb{E}\varphi(\bar{\mathbf{x}}^*, \boldsymbol{\xi}) \\ &= \sum_{t=2}^T \mathbb{E} \left[\langle b_t(\tilde{\boldsymbol{\xi}}_t) - b_t(\boldsymbol{\xi}_t), \bar{\mathbf{x}}_t^*(\tilde{\boldsymbol{\xi}}^t) \rangle + \langle b_t(\boldsymbol{\xi}_t), \bar{\mathbf{x}}_t^*(\tilde{\boldsymbol{\xi}}^t) - \bar{\mathbf{x}}_t^*(\boldsymbol{\xi}^t) \rangle \right] \\ &\leq B \sum_{t=2}^T \mathbb{E} \left[\|\tilde{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t\| \|\bar{\mathbf{x}}_t^*(\tilde{\boldsymbol{\xi}}^t)\| + \max\{1, \|\boldsymbol{\xi}_t\|\} \|\bar{\mathbf{x}}_t^*(\tilde{\boldsymbol{\xi}}^t) - \bar{\mathbf{x}}_t^*(\boldsymbol{\xi}^t)\| \right]. \end{aligned}$$

Each of these $T - 1$ summands can be estimated by boundedness of \mathbf{x}^* and calmness of $\bar{\mathbf{x}}^*$. Thus, (25) is bounded by

$$(26) \quad H\mathbb{E} \left[\max\{1, \|\boldsymbol{\xi}^T\|, \|\tilde{\boldsymbol{\xi}}^T\|\}^T \cdot \|\boldsymbol{\xi}^T - \tilde{\boldsymbol{\xi}}^T\| \right],$$

with an appropriate constant $H > 0$. Hence, using again $m_1 \geq T$, we obtain

$$v(\tilde{\boldsymbol{\xi}}) - v(\boldsymbol{\xi}) \leq (C + H)\mathbb{E} \left[\max\{1, \|\boldsymbol{\xi}^T\|, \|\tilde{\boldsymbol{\xi}}^T\|\}^{m_1} \cdot \|\boldsymbol{\xi}^T - \tilde{\boldsymbol{\xi}}^T\| \right].$$

Now, we consider a bounded optimal decision $\tilde{\mathbf{x}}^*$ of $v(\tilde{\boldsymbol{\xi}})$. Following exactly the construction of Section 3, we obtain a decision $\bar{\tilde{\mathbf{x}}}^* \in \mathcal{S}(\boldsymbol{\xi})$ that is calm in the sense of Proposition 3.1 and whose restriction on $f^T(\Xi^T)$ is optimal for $v(\tilde{\boldsymbol{\xi}})$. As in (25), it follows that

$$\begin{aligned} v(\boldsymbol{\xi}) - v(\tilde{\boldsymbol{\xi}}) &\leq \mathbb{E}\varphi(\bar{\tilde{\mathbf{x}}}^*, \boldsymbol{\xi}) - \mathbb{E}\varphi(\tilde{\mathbf{x}}^*, \tilde{\boldsymbol{\xi}}) \\ &\leq B \sum_{t=2}^T \mathbb{E} \left[\max\{1, \|\boldsymbol{\xi}_t\|\} \|\bar{\tilde{\mathbf{x}}}_t^*(\tilde{\boldsymbol{\xi}}^t) - \tilde{\mathbf{x}}_t^*(\boldsymbol{\xi}^t)\| + \|\tilde{\boldsymbol{\xi}}_t - \boldsymbol{\xi}_t\| \|\tilde{\mathbf{x}}_t^*(\boldsymbol{\xi}^t)\| \right] \\ &\leq H\mathbb{E} \left[\max\{1, \|\boldsymbol{\xi}^T\|, \|\tilde{\boldsymbol{\xi}}^T\|\}^T \cdot \|\boldsymbol{\xi}^T - \tilde{\boldsymbol{\xi}}^T\| \right]. \end{aligned}$$

The proof is completed by setting $\gamma \triangleq C + H$. □

Remark 4.1. *Since the scope of this paper is rather to establish a stability result than the development of new approximation techniques, we restrict ourselves to refer to existing approaches based on conditional or unconditional clustering, that can be used to control the upper bound of Theorem 3. We mention here the approaches of Heitsch and Römisch (2005), Bally, Pagès, and Printems (2005), Hochreiter and Pflug (2007), and Pennanen (2007).*

Appendix

The following lemma provides conditions under which the conditions of Assumption 2.6 hold true.

Lemma A.1. *Assume the dynamic of the process ξ is given by the following scheme:*

$$(27) \quad \xi_{t+1} = g_t(\xi^t, \varepsilon_{t+1}),$$

where ε_{t+1} is a random variable that is independent of ξ^t and g_t are measurable mappings from $\mathbb{R}^{m_t} \times \mathbb{R}^s$ to \mathbb{R}^m that satisfy the following Lipschitz and linear growth conditions:

$$(i) \quad \|g_t(x, \varepsilon_{t+1}) - g_t(y, \varepsilon_{t+1})\| \leq \max\{1, \|x\|, \|y\|\}^r \|x - y\| \cdot h(\|\varepsilon_{t+1}\|),$$

$$(ii) \quad \|g_t(x, \varepsilon_{t+1})\| \leq \max\{1, \|x\|\} \cdot k(\|\varepsilon_{t+1}\|),$$

for an $r \geq 1$ and Borel-measurable mappings $h, k \geq 1$, such that $h(\|\varepsilon_{t+1}\|)$ and $k(\|\varepsilon_{t+1}\|)$ are in L^p for every $p \in [1, +\infty)$. Then ξ fulfils both conditions of Assumption 2.6 with the constants

$$K \triangleq \mathbb{E} [k(\|\varepsilon_{t+1}\|)^{m_1} h(\|\varepsilon_{t+1}\|)] \quad \text{and} \quad W \triangleq \mathbb{E} \left[\prod_{i=t+1}^T k(\|\varepsilon_i\|)^{2+T-t} \right].$$

Proof. Consider $f \in \mathbb{F}_{m_{t+1}}(\Xi^{t+1})$. Then we obtain

$$\begin{aligned} & \left| \mathbb{E} [f(\xi^{t+1}) | \xi^t = \xi^t] - \mathbb{E} [f(\xi^{t+1}) | \xi^t = \hat{\xi}^t] \right| \\ &= \left| \mathbb{E} [f(g_t(\xi^t, \varepsilon_{t+1}))] - \mathbb{E} [f(g_t(\hat{\xi}^t, \varepsilon_{t+1}))] \right| \\ &\leq \mathbb{E} \left[\max \left\{ 1, \|g_t(\xi^t, \varepsilon_{t+1})\|, \|g_t(\hat{\xi}^t, \varepsilon_{t+1})\| \right\}^{m_{t+1}} \|g_t(\xi^t, \varepsilon_{t+1}) - g_t(\hat{\xi}^t, \varepsilon_{t+1})\| \right] \\ &\leq \mathbb{E} \left[\max \left\{ 1, \|g_t(\xi^t, \varepsilon_{t+1})\|, \|g_t(\hat{\xi}^t, \varepsilon_{t+1})\| \right\}^{m_{t+1}} h(\|\varepsilon_{t+1}\|) \right] \\ &\quad \cdot \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^r \|\xi^t - \hat{\xi}^t\| \\ &\leq \mathbb{E} \left[\max \left\{ 1, \|\xi^t\|, \|\hat{\xi}^t\| \right\}^{m_{t+1}} k(\|\varepsilon_{t+1}\|)^{m_{t+1}} h(\|\varepsilon_{t+1}\|) \right] \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^r \|\xi^t - \hat{\xi}^t\| \\ &= \mathbb{E} [k(\|\varepsilon_{t+1}\|)^{m_{t+1}} h(\|\varepsilon_{t+1}\|)] \cdot \max\{1, \|\xi^t\|, \|\hat{\xi}^t\|\}^{r+m_{t+1}} \|\xi^t - \hat{\xi}^t\|. \end{aligned}$$

Due to the identity $r + m_{t+1} = m_t - 1$, this entails condition (i) of Assumption 2.6. The asserted form of K follows from $m_1 \geq m_t$ for $t = 1, \dots, T$.

Furthermore, we apply (27) recursively to obtain the following estimate:

$$\|\xi^T\| \leq \max\{1, \|\xi^t\|\} \prod_{i=t+1}^T k(\|\varepsilon_i\|).$$

Raising both sides to the power of $2 + (T - t)$ and taking conditional expectations

$\mathbb{E}[\cdot | \xi^t = \xi^t]$ verifies condition (ii) of Assumption 2.6. \square

The conditions of Lemma A.1 are fulfilled, e.g., by a variety of time-series models. We give the following simple

Example A.2. Let ξ be a GARCH process defined through the following difference equations:

$$\xi_t = (\mathbf{w}_t, \mathbf{v}_t, \varepsilon_t) \text{ with}$$

$$\mathbf{v}_{t+1} \triangleq \sum_{i=0}^s (\beta_i v_{t-i} + \gamma_i \varepsilon_{t-i}) \quad \text{and} \quad \mathbf{w}_{t+1} \triangleq \sum_{i=0}^s \alpha_i \mathbf{w}_{t-i} + \mathbf{v}_{t+1} \cdot \varepsilon_{t+1}$$

for certain parameters $\alpha_i, \beta_i, \gamma_i \in \mathbb{R}$. Thereby, \mathbf{v} represents the stochastic volatility process of \mathbf{w} and $(\varepsilon_t)_{t \geq 0}$ is a sequence of i.i.d. random variables, following a standard normal distribution. It is easy to see that ξ fulfills the conditions of Lemma A.1 with $r = 1$ and $h(\cdot), k(\cdot)$ being affine functions.

The following example shows that nonanticipativity w.r.t. the initial process is indispensable for an approximating process.

Example A.3. Consider $T = 3$ and the process ξ that is given by $\xi_1 \equiv 0$ and the two independent random variables ξ_2 and ξ_3 , both uniformly distributed on $[0, 1]$. For $n \in \mathbb{N}$ and $0 < \varepsilon < 1$ we introduce the grids $A^{(n)} \triangleq \{\frac{i}{n} : i = 1, \dots, n\}$ and the associated projections $\pi_{A^{(n)}} : [0, 1] \rightarrow A^{(n)}$.

Furthermore, we define processes $\xi^{(n)}, n \in \mathbb{N}$, given by $\xi_1^{(n)} \equiv 0$, $\xi_3^{(n)} \triangleq \pi_{A^{(n)}} \xi_3$, and

$$\xi_2^{(n)} \triangleq \begin{cases} \pi_{A^{(n)}} \xi_2 & \text{if } \xi_3 \leq 1/2, \\ (\pi_{A^{(n)}} \xi_2) + \frac{\varepsilon}{n} & \text{if } \xi_3 > 1/2. \end{cases}$$

The sequence $\xi^{(n)}$ can be seen as an approximation of ξ , since $\mathbb{E} [\|\xi - \xi^{(n)}\|] \leq \frac{1+2\varepsilon}{2n}$ holds true. We consider the following optimization problem

$$v(\xi) \triangleq \min \{ \mathbb{E} [\mathbf{x}_2 \cdot \xi_2 + \mathbf{x}_3 \cdot \xi_3] : \mathbf{x}_t \geq 0, \mathbf{x}_t \in \sigma(\xi^t), t = 2, 3, \mathbf{x}_2 + \mathbf{x}_3 = 1 \text{ a.s.} \},$$

that is solved by $\mathbf{x}_2 = \mathbb{1}_{\{\xi_2 \leq 1/2\}}$ and $\mathbf{x}_3 = 1 - \mathbf{x}_2$ with optimal value $v(\xi) = 12/32$. When replacing ξ by $\xi^{(n)}$, we use the decisions

$$\mathbf{x}_2^{(n)} = \mathbb{1}_{\{\xi_2^{(n)} \leq 1/4\}} + \mathbb{1}_{\{\xi_2^{(n)} \in]1/4, 3/4[\setminus A^{(n)}\}} \quad \text{and} \quad \mathbf{x}_3^{(n)} = 1 - \mathbf{x}_2^{(n)}$$

to obtain $\limsup_{n \rightarrow \infty} v(\xi^{(n)}) \leq 11/32$. Obviously, convergence does not hold since the processes $\xi^{(n)}$ do not fulfill the nonanticipativity condition (i) of Definition 2.8.

References

- BALLY, V., G. PAGÈS, AND J. PRINTEMPS (2005): “A quantization tree method for pricing and hedging multidimensional american options,” *Mathematical Finance*, 15(1), 119–168.
- BARTY, K. (2004): “Contributions à la discrétisation des contraintes de mesurabilité pour les problèmes d’optimisation stochastique,” Thèse de Doctorat, École Nationale des Ponts et Chaussées, Paris.
- BIRGE, J., AND F. LOUVEAUX (1997): *Introduction to Stochastic Programming*, Springer Series in Operations Research. Springer-Verlag.
- DOBRUSHIN, R. (1956): “Central limit theorem for non-stationary Markov chains I.,” *Teor. Veroyatnost. i Primenen.*, 1(N1), 72–89.
- DYNKIN, E. (1965): *Markov processes*. Springer, Berlin.
- EVSTIGNEEV, I. (1976): “Measurable selection and dynamic programming,” *Mathematics of Operations Research*, 1(3), 267–272.
- HEITSCH, H., AND W. RÖMISCH (2005): “Scenario tree modelling for multistage stochastic programs,” Preprint 296, DFG Research Center Matheon ”Mathematics for key technologies” and submitted.
- HEITSCH, H., W. RÖMISCH, AND C. STRUGAREK (2006): “Stability of multistage stochastic programs,” *SIAM Journal on Optimization*, 17, 511–525.
- HOCHREITER, R., AND G. PFLUG (2007): “Financial scenario generation for stochastic multi-stage decision processes as facility location problems,” *Annals of Operations Research*, 152(1), 257–272.
- KUHN, D. (2005): *Generalized bounds for convex multistage stochastic programs*, vol. 548 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin.
- MIRKOV, R., AND G. PFLUG (2006): “Tree Approximations of Dynamic Stochastic Programs,” *submitted*.
- PENNANEN, T. (2005): “Epi-convergent Discretizations of Multistage Stochastic Programs,” *Mathematics of Operations Research*, 30(1), 245–256.
- (2007): “Epi-convergent discretizations of multistage stochastic programs via integration quadratures,” *Mathematical Programming*, to appear.
- RACHEV, S. (1991): *Probability Metrics and the Stability of Stochastic Models*. Wiley.

- ROCKAFELLAR, R., AND R. J.-B. WETS (1974): “Continuous Versus Measurable Recourse in N -Stage Stochastic Programming,” *Journal of Mathematical Analysis and Applications*, 48(3), 836–859.
- (1998): *Variational Analysis*. Springer-Verlag, Berlin.
- RÖMISCH, W. (2003): *Stability of Stochastic Programming Problems*, chap. 8, pp. 483–554, in Ruszczyński and Shapiro (2003).
- RUSZCZYŃSKI, A., AND A. SHAPIRO (eds.) (2003): *Stochastic Programming*, Handbooks in Operations Research and Management Science. Elsevier, Amsterdam.
- SHAPIRO, A. (2003): “Inference of statistical bounds for multistage stochastic programming problems,” *Mathematical Methods of Operations Research*, 58(1), 57–68.